



Briefings in Bioinformatics, 17(3), 2016, 527–540

doi: 10.1093/bib/bbv062

Advance Access Publication Date: 24 August 2015

Paper

Systems medicine of inflammaging

Gastone C. Castellani*, Giulia Menichetti*, Paolo Garagnani, Maria Giulia Bacalini, Chiara Pirazzini, Claudio Franceschi, Sebastiano Collino, Claudia Sala, Daniel Remondini, Enrico Giampieri, Ettore Mosca, Matteo Bersanelli, Silvia Vitali, Italo Faria do Valle, Pietro Liò and Luciano Milanese

Corresponding author: Gastone C. Castellani. Department of Physics and Astronomy viale Berti Pichat 6/2, 40127, Bologna, Italy. E-mail: gastone.castellani@unibo.it

*These two authors equally contributed to this paper.

Abstract

Systems Medicine (SM) can be defined as an extension of Systems Biology (SB) to Clinical-Epidemiological disciplines through a shifting paradigm, starting from a cellular, toward a patient centered framework. According to this vision, the

Gastone C. Castellani is Associate Professor (Bologna University Physics Department), director of the Galvani center for Biocomplexity. Associate Professor at Brown University and author of more than 200 peer-reviewed papers.

Giulia Menichetti (PhD in Physics) is Postdoc in the Biophysics group (Bologna University Physics Department). Her interests include network theory, multilayer networks, data analysis applied to Systems Biology.

Paolo Garagnani (PhD) is Research assistant (Bologna University, Experimental, Diagnostic and Specialty Medicine Department). His interests include human genetics and epigenetics, applied to aging and longevity phenotypes.

Maria Giulia Bacalini (PhD) is Post-Doc (Bologna University, Experimental, Diagnostic and Specialty Medicine Dept). Her current research is focused on the identification of DNA methylation alterations in age-related diseases.

Chiara Pirazzini (PhD) is Post-Doc (Bologna University, Experimental, Diagnostic and Specialty Medicine Dept). Her research activity is focused on genetic and epigenetic basis of human longevity.

Claudio Franceschi is Full Professor (Bologna University, Experimental, Diagnostic and Specialty Medicine Department). He was the first to conceptualize the theory of inflammaging and he contributed to more than 500 peer-reviewed papers.

Sebastiano Collino (PhD) is expert in the fields of metabolomics, biomarker discovery and nutrition. He is currently working at NIHS. Sebastiano's research focuses on systems biology approaches to study healthy aging.

Claudia Sala is PhD student in the Biophysics group (Bologna University Physics Department). Her main research activities include stochastic modeling and data analysis applied to Gut Microbiota.

Daniel Remondini is Associate Professor (Bologna University Physics Department). His research interests are in the field of Biophysics and network theory applied to Systems Biology and is author of more than 100 peer-reviewed papers.

Enrico Giampieri (PhD in Physics) is Postdoc in the Biophysics group (Bologna University Physics Department). His research involves stochastic processes, data analysis and Bayesian Statistics applied to Systems Biology.

Ettore Mosca is Research Assistant at the ITB-CNR. His research activities include: data integration and databases, modeling of biological systems, analysis (parameter estimation, sensitivity analysis) of dynamical systems.

Matteo Bersanelli is PhD student in the Biophysics group (Bologna University Physics Department and ITB-CNR). His research activities include mathematical stochastic modeling, applied to Systems Medicine.

Silvia Vitali is PhD student in the Biophysics group (Bologna University Physics Department). Her main research activities include stochastic modeling and data analysis of DNA transposable elements.

Italo Faria do Valle is PhD student (Bologna University Biochemistry and Biotechnology Department). His research activity is on the analysis of Next-Generation Sequencing data.

Pietro Liò is reader in Computational Biology (Cambridge University, Computer Science Department). He is interested in Machine Learning and computational models in health Big Data. He is author in more than 300 publications.

Luciano Milanese is coordinator of the CNR Interdepartmental Bioinformatics Research Network in Life science, Medicine and ICT. He is author in more than 300 publications.

Submitted: 23 April 2015; Received (in revised form): 29 June 2015

© The Author 2015. Published by Oxford University Press.

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

three pillars of SM are Biomedical hypotheses, experimental data, mainly achieved by Omics technologies and tailored computational, statistical and modeling tools. The three SM pillars are highly interconnected, and their balancing is crucial. Despite the great technological progresses producing huge amount of data (Big Data) and impressive computational facilities, the Bio-Medical hypotheses are still of primary importance. A paradigmatic example of unifying Bio-Medical theory is the concept of Inflammaging. This complex phenotype is involved in a large number of pathologies and patho-physiological processes such as aging, age-related diseases and cancer, all sharing a common inflammatory pathogenesis. This Biomedical hypothesis can be mapped into an ecological perspective capable to describe by quantitative and predictive models some experimentally observed features, such as microenvironment, niche partitioning and phenotype propagation. In this article we show how this idea can be supported by computational methods useful to successfully integrate, analyze and model large data sets, combining cross-sectional and longitudinal information on clinical, environmental and omics data of healthy subjects and patients to provide new multidimensional biomarkers capable of distinguishing between different pathological conditions, e.g. healthy versus unhealthy state, physiological versus pathological aging.

Key words: propagation; ecological model; networks; multilayer networks; inflammation; multi-scale

Introduction

There is a large debate about a possible definition of Systems Medicine (SM). One idea is to define SM as an implementation of Systems Biology (SB) in the Medical disciplines with a particular attention to clinical applications [1], including clinical Bioinformatics and the discrimination of pathological states and related morbidities and comorbidities.

The extension of SB (Table 1) to the clinical practice implies the establishment of a connection between a molecular-centered to a patient-centered world, through an organ-centered intermediate layer. This mapping (Figure 1) requires the extensive use of computational tools such as statistical, mathematical and bioinformatical techniques [1].

Within this perspective, SM is deeply related to Personalized (PeM) and Precision Medicine (PrM) [1–3]. PeM and PrM share the idea of the centrality of the individual patient as the common denominator for the development of tailored approaches such as therapies, drugs and treatments, according to the genetic background and, more in general, to the individual microenvironment [3].

The role of the microenvironment is becoming crucial in the comprehension of a variety of pathological and patho-physiological processes, such as cancer, diabetes, aging and age-related diseases [4–6]. Indeed, SM exploits, characterizes and quantifies the concept of microenvironment by novel analytical methods based on omics technologies and by the development of adequate computational models [7].

A research field where the microenvironment is playing a central role is those of aging and age-related diseases. The

microenvironment, mainly the circulating one, has been described as related to inflammation, contributing to the definition of inflammaging, a complex phenotype that impacts in a variety of age-related diseases [4–6] as demonstrated by a series of experiments, collectively called ‘parabiosis’ [8]. These results gave birth to the concept of Communicome, defined as the set of plasmatic proteins involved in the rejuvenation effect [8].

The parabiosis experiments can be interpreted within the framework of the so-called ‘bystander effect’ [9]. The bystander effect has been observed during aging progression, cancer induction and other pathologies in which it is believed to explain the systemic propagation between cells and organs [10].

In the case of human body, the most accessible fluids are the serum and urines. The composition of these fluids can be assessed by a variety of methods, aimed to identify different fractions, such as metabolite concentrations, proteins and nucleic acids. The identification of these fractions is accomplished with omics methodologies such as metabolomics, proteomics and deep sequencing. Thus the body fluids, including both extra- and intracellular, belong to the individual microenvironment (‘le milieu interior’) [11]. Our microenvironment is in dynamical equilibrium with a variety of cells, hosting, among others, nutrients and molecular messages crucially involved in the regulation of the cellular populations. The microenvironment can be considered as a particular ‘habitat’ where different cells, and in some cases other hosted organisms, live and interact. Within this perspective, the role of microenvironment in shaping the other components of the body can be modeled from an ecological point of view. The same ecological approach can be used

Table 1. Increasing complexity and spreading of inflammation

Layer	BioMed hypothesis	Technology	Computation
Cellular	Inflammation Senescence	Epigenetic dysregulation Microscopy	Stochastic modeling Statistical distribution analysis inferential testing
Organ-tissue	Spreading of inflammation, including propagation of senescence	Sequencing Metabolomics Metagenomics	Bayesian methods Ecological modeling Diffusion models
Patient	Systemic inflammation Cancer Aging	Imaging (PET) Deep sequencing, including SNP detection	Population models Association testing (odds ratio, logistic regression) Texture analysis

Note. The propagation of inflammation through different layers of complexity can be disentangled by ad hoc combination of computational and technological tools, aimed to validate the biomedical hypotheses.

for the modeling of ecosystems that are in parabiosis with our body such as the gut microbiota (GM), the circulating virome (CV) and the transposable elements (TE) in our DNA [12–14], which have coevolved with the human host and his immune system (IS). This evolutionary-ecological framework is deeply related to inflammaging progression, given the role of the microenvironment, whose composition modulates the inflammatory process (pro- or anti-inflammatory action).

As a further computational tool, we have to mention complex networks [15] and their extension to multiple networks: multiplex and multilayer networks [16, 17]. Multilayer networks can be used for the representation and quantification of the interactions arising from the internal environment of the human body and for those coming from the external environment. The architecture of complex networks is a natural embedding for diffusion processes as a function of topological constraints such as those observed during inflammatory processes.

The unifying biomedical hypothesis: inflammaging as the result of communicome and propagation

Chronic, low-grade inflammation is recognized as a main characteristic of aging bodies. This phenomenon, also indicated as inflammaging [4] is particularly important because practically all the major age-related diseases have an inflammatory pathogenic component [5]. Thus, age is a major risk factor for these diseases, and because they all share inflammaging as a prominent determinant, it is clear that inflammaging is a privileged target to combat age-related diseases as a whole and not one by one [6]. According to the most recent studies, inflammaging appears to be much more complex and entangled than we previously thought, and a variety of interconnected tissues, organs and systems participate in producing inflammatory stimuli [4–6].

Among them we mention the IS, the adipose tissue, the muscles and the liver. A particular attention deserves the GM,

and the variety of inflammatory stimuli sent through the general circulation to all the body organs [18].

The general idea is that the inflammatory stimuli fueling inflammaging can be exogenous (e.g. persistent viral infections, diet, lifestyle, see Figure 2) [19], but probably most of them are endogenously produced by the body and should be identified with the self-debris resulting from the continuous turnover of cells and tissues [4], e.g. circulating DNA and RNA (the virome) [13], pro-inflammatory agalactosylated N-glycans [20] and pro-inflammatory circulating microRNA ('inflammaMIR') [21].

There is a general consensus that the most important phenomena involved in inflammaging at the cellular and molecular level are the following:

(i) Cell senescence and its pro-inflammatory senescent-associated secretory phenotype triggered by damaging agents (radiation, viruses) and likely by the continuous exposure to the above-mentioned self-debris [22]. It is interesting that cell senescence can spread to neighboring cells ('senescence by senescence') [23].

(ii) DNA damage, including telomeres, by Reactive Oxygen Species (ROS) and by a variety of other agents, which in turn triggers a DNA damage response and the production of pro-inflammatory compounds [24].

All these mechanisms suggest targeted therapies aimed to reduce inflammaging: elimination of senescent cells [25], diet enriched with omega3 fatty acids, Mediterranean diet [26] and other nutritional strategies, including caloric restriction (adapted to humans and taking into account genetic background) and intermittent fasting (Figure 2).

Contemporary research is highlighting the role of inflammation in cancer induction and progression. Inflammaging seems therefore a unifying mechanism shared by different pathological processes and not just by aging and age-related diseases. The role of inflammation for cancer progression has been hypothesized by ancient Greeks and progressively confirmed by a number of researchers [27]. The inflammatory state is sustained by cells and molecules that shape the microenvironment

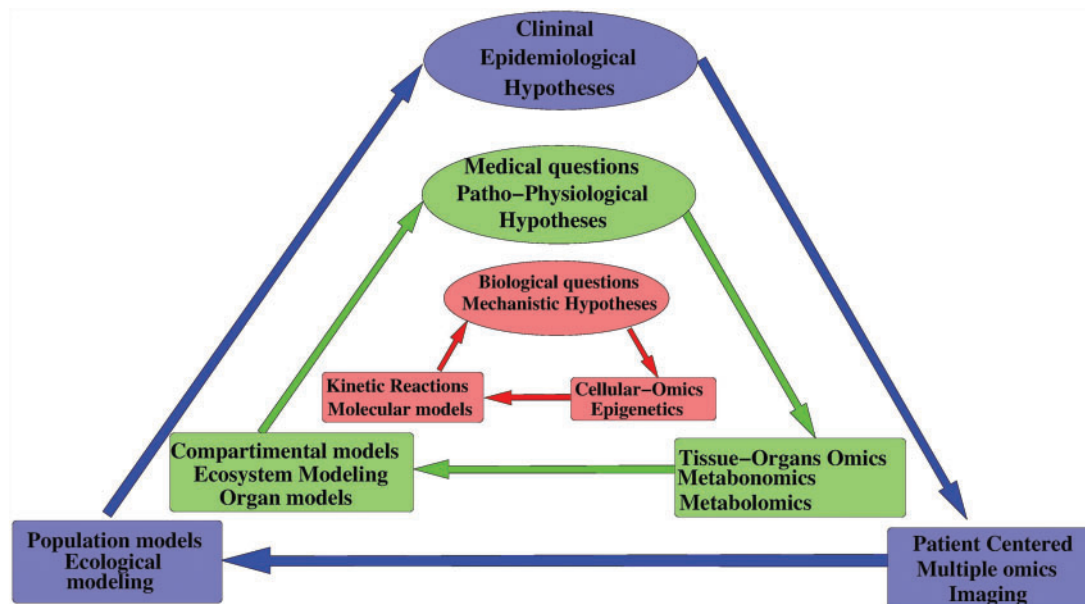


Figure 1. SM as extension of SB. The SB basic cycle (red) is composed of Biological hypotheses, Technology (mainly devoted to omics measurements) and computational tools (statistical and modeling methods). The core SB cycle is then extended by increasing its complexity in a multi-scale way; starting from the cellular-subcellular domain, we reach the individual domain (the domain identified by a single patient and its internal structure in terms of organs and tissues). Finally, the higher domain is constituted by collection of patients and their interactions in an epidemiological context.

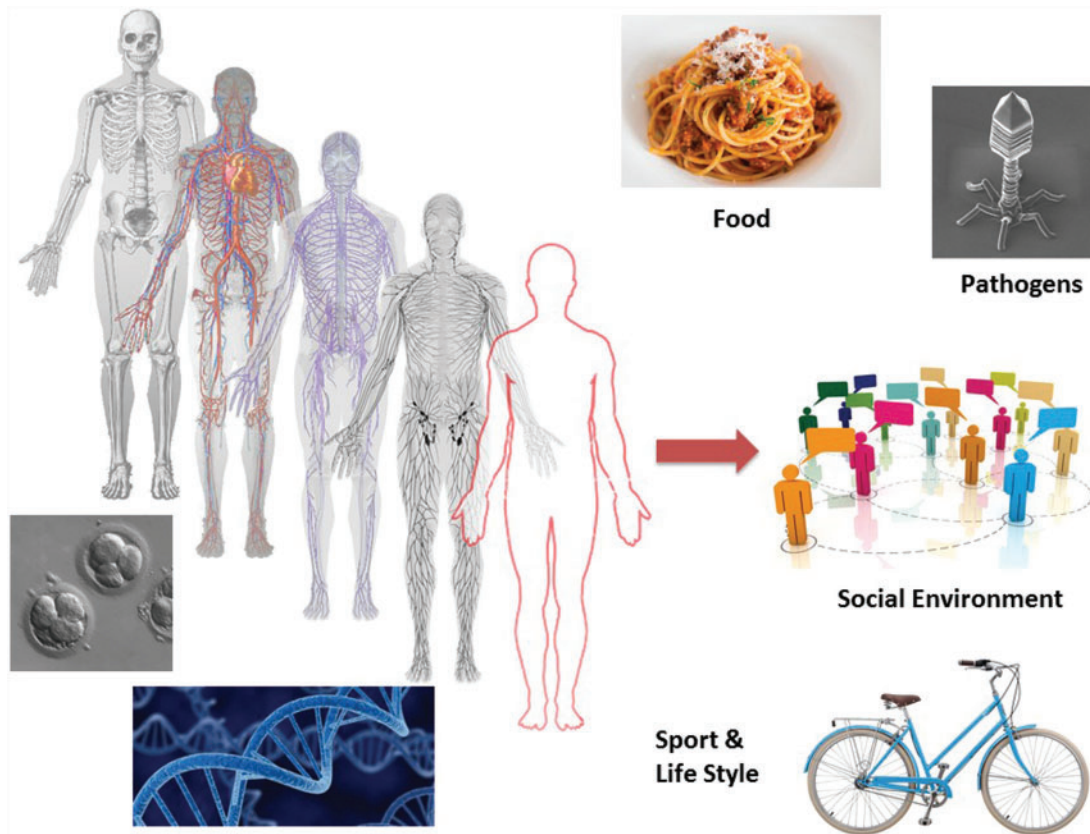


Figure 2. Patient data space. Personalized medicine gathers together a huge amount of data characterizing a single patient (red shape). This integrated system can be interpreted as a network of networks. These integrated networks belong to different complexity layers, i.e. the omics layer (e.g. DNA and cells), the anatomical-functional layer (e.g. skeleton, circulatory system, nervous system, lymphatic system) and finally, the environmental layer (e.g. dietary habits, diseases and drugs, social behaviors and sports). In particular, the anatomical-functional scale is a spatial multiplex network, i.e. nodes can be considered as specific regions in space, connected by different functional links. A colour version of this figure is available at BIB online: <http://bib.oxfordjournals.org>.

(Figure 3). Chronic inflammation affects all cancer stages: it increases the onset risk, supporting the initial genetic mutation or epigenetic mechanism, and it leads to cancer initiation, promoting tumor progression and supporting metastatic diffusion.

The spreading character of inflammation has been quantified by the concept of Communicome, defined by [8] as the circulating plasma proteins (i.e. microenvironment) with endocrine activity, including inflammatory cytokines, hormone-like proteins, growth factors and so forth. These factors are involved in intercellular and organ communication, and their composition changes with age and pathologies are crucial for discriminating normal versus pathological trajectories.

Another major player in shaping the individual microenvironment is the GM (Figure 3). GM is a microbial ecosystem with a complex dynamics that derives from the interactions with components such as the virome (the set of viruses in the host body) and the IS [28]. The collective genome of these symbiotic microorganisms (called 'Microbiome') [29] constantly interacts with the host genome, forming the so called 'Metagenome' [30]. The microbial genome is in dynamical relation with the host organism, helping in crucial functions such as metabolic processes (e.g. food absorption), short-chain fatty acids and vitamin production [31], but also shaping, controlling and protecting the IS development [32], so fueling the (co)-evolution of the host.

It is through the interaction between the different components of the metagenome that the host health phenotype is defined [33]. GM is linked, through an interdependence relationship, to the host IS [34] and metabolism [35], becoming crucial

for a large number of physio-pathological conditions and diseases. These include inflammatory and metabolic diseases, in particular obesity [36], metabolic syndrome (MS) and type 2 diabetes, as well as aging [37] and cancer. Nowadays, the availability of Next-Generation Sequencing (NGS) Methods, for the characterization of bacterial communities, contributed to the creation of a new research field, called Metagenomics. Metagenomics is the set of omics measurements that quantify the composition of the metagenome and the interactions between the host and the microbiome at multiple levels: DNA (metagenome), RNA (meta-transcriptome), protein (meta-proteome) and metabolic network (metabolome).

The ecological framework: the examples of GM and TE

Two relevant examples of the microenvironment-related ecological hypothesis feasibility are GM and TEs.

Regarding GM, there is a growing interest to the quantification of biodiversity and dynamics that brings to a certain composition of this complex ecosystem and, moreover, for the assessment of its homeostasis degree, its support capacity, as well as for the prediction of its temporal evolution. From this perspective it is clear that mathematical models have a major role [38]. GM dynamics can be described from an ecological point of view, with a focus on GM biodiversity, and in particular on one of its components: the Relative Species Abundance distribution (RSA). RSA is defined within a single phylogenetic level

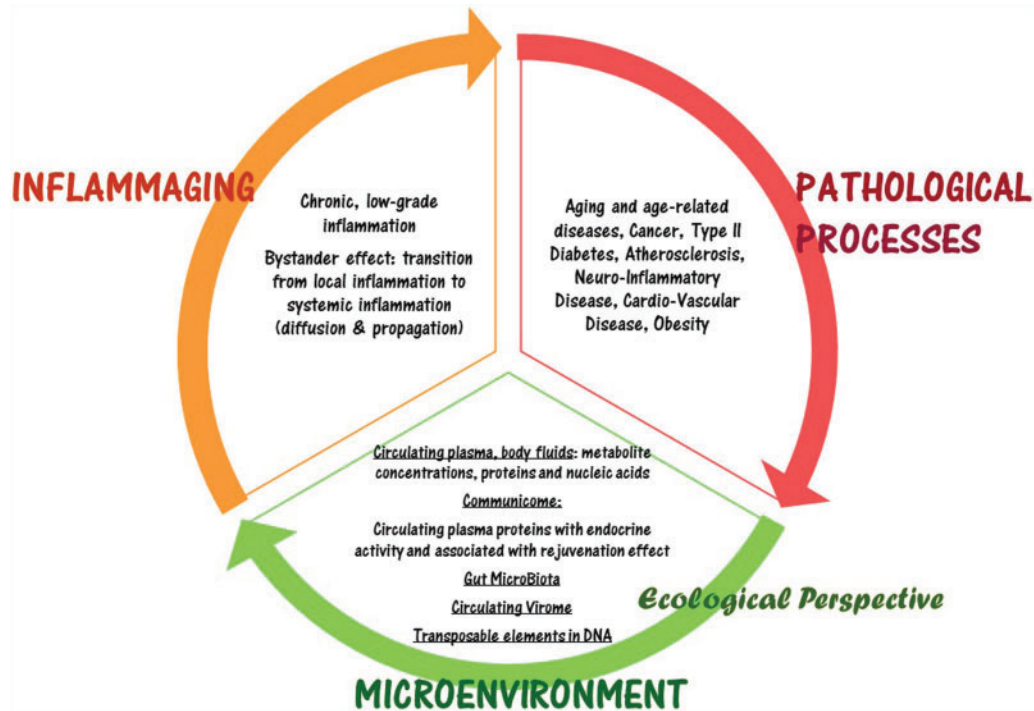


Figure 3. The unifying BioMedical hypothesis of Inflammation as a driver of inflammaging and age-related pathological processes. The propagating (bystander) character of inflammation and inflammaging is strictly related to the microenvironment, including the circulating one (e.g. the communicome, virome, mobilome and GM), and its suggested interpretation from an ecological point of view.

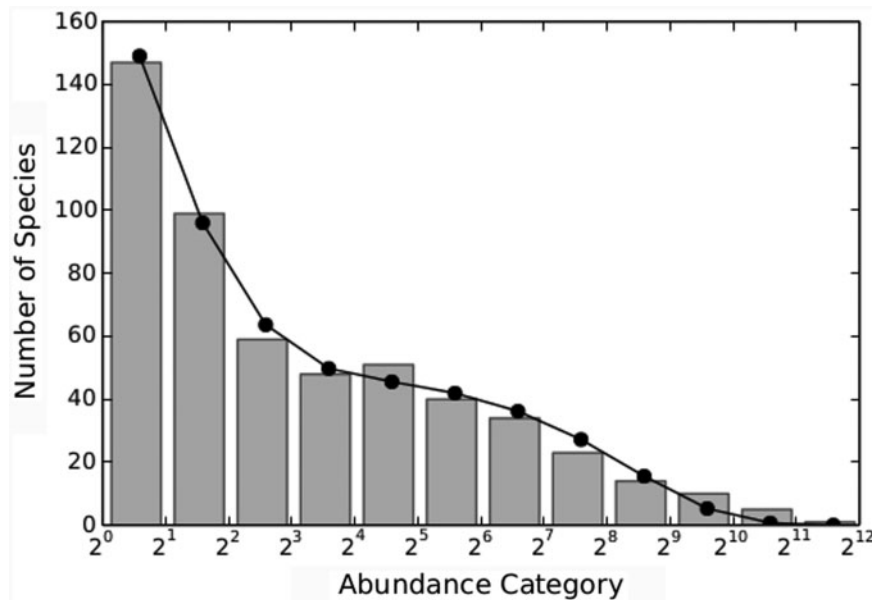


Figure 4. RSA of one GM sample from Claesson *et al.* (gray histogram) fitted with a mixture of two Negative Binomials (black line). The RSA is a measure of biodiversity and is usually represented in the form of Preston Plot, plotting the number of species that have a certain number of individuals (in log 2). The neutral model proposed by Volkov *et al.* predicts a Negative Binomial distribution for the RSA and fits well the TE population. The GM RSA is rather fitted by a mixture of two Negative Binomials, meaning that a relaxation of the neutrality assumption is needed. The GM is thus well described by a hybrid niche-neutral model, in which two neutral niches are considered.

and refers to how common or rare a species is in comparison with other species and it is derived from a single family of statistical distributions, ranging from the Log-Series to a highly skewed and unveiled Log-Normal [39] (Figure 4). Modern ecological theories can be distinguished in essentially two main schools of thought: the niche assembly perspective and the dispersal one [40].

The niche assembly perspective states that communities are groups of interacting species whose presence or absence and even their relative abundance can be deduced from deterministic ‘assembly rules’ that are based on the ecological niches or functional roles of each species. According to this view, species coexist in an interactive balance and a stable coexistence among competing species is made possible by niche partitioning [41].

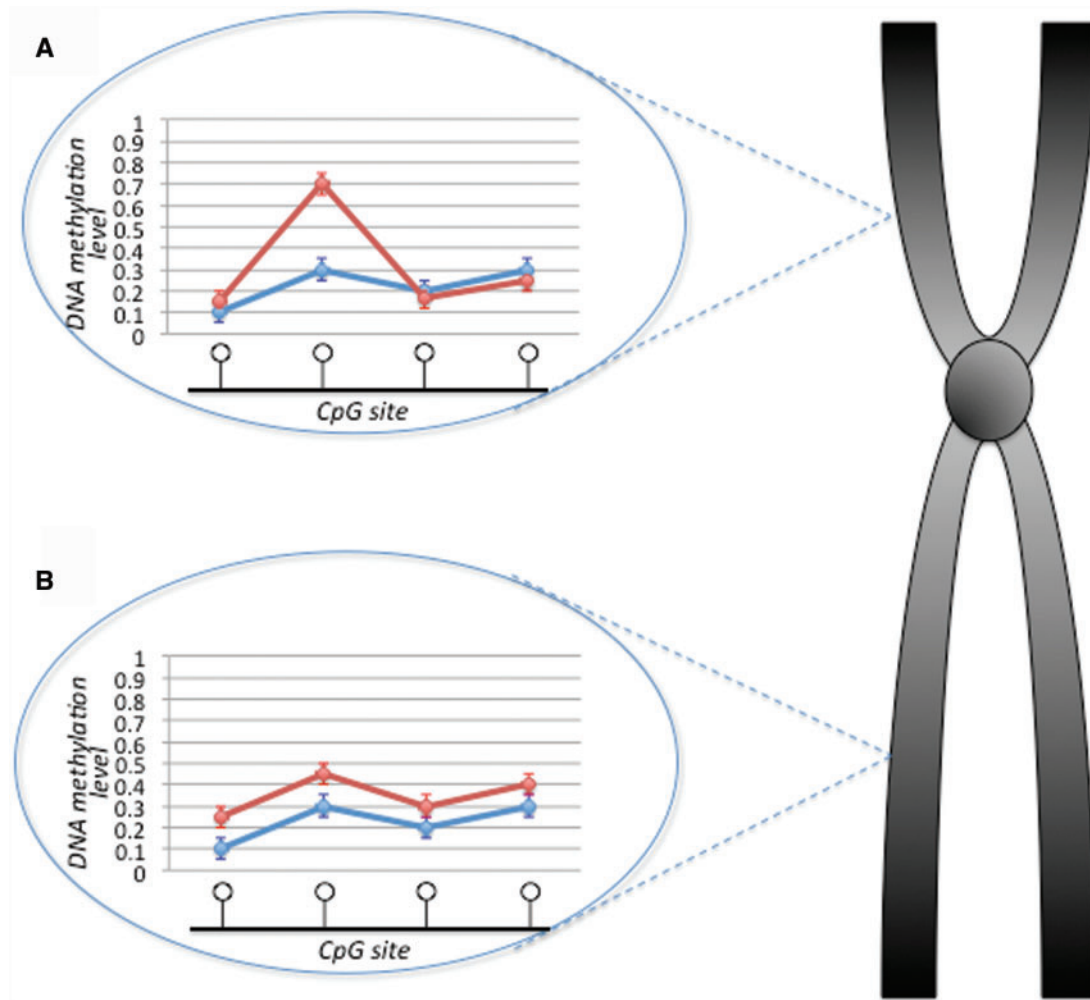


Figure 5. Single-probe versus region-centric approaches for the analysis of DNA methylation microarrays. CpG DNA methylation measured by Infinium 450 k microarrays is expressed as a continuous value ranging from 0 (the CpG site is unmethylated in all the analyzed DNA molecules) to 1 (the CpG site is unmethylated in all the analyzed DNA molecules). In the analysis of differential methylation between two groups of samples, single-probe and region-centric approaches return different results. Single-probe analysis favors genomic regions like the one reported in box A, where a unique CpG site strongly differs in its methylation value between group A and group B. In the genomic region reported in box B, on the contrary, differences in DNA methylation values between groups A and B are smaller, but they involve several adjacent CpG sites; this configuration is preferentially identified by a region-centric approach. In the figure, CpG sites are represented as lollipops. A colour version of this figure is available at BIB online: <http://bib.oxfordjournals.org>.

On the other hand, the dispersal assembly (neutral theory) perspective asserts that communities are open, non-equilibrium assemblages of species largely put together by chance, history and random dispersal [40]. Species come and go, and their presence or absence is dictated by random dispersal and stochastic local extinction. Ecological communities are structured entirely by ecological drift (i.e. demographic stochasticity), random migration and random speciation.

These two theories are not mutually exclusive; in fact it is possible to identify regimes where both of them are present. In particular, inflammaging supports both theories: the micro-environment concept is conceivable as an ecological niche, while the neutral theory efficiently describes and predicts RSA shape also in presence of external perturbations (diet, lifestyle, etc.) both for GM and TE.

TE ecosystem (the so-called mobilome) is another component of the microenvironment, related to the CV and describable as an ecological system living in parabiosis or symbiosis (sometimes parasitism) relationship with the host DNA. TEs are also known as selfish DNA or jumping genes. They are present in DNA of

eukaryotic and prokaryotic organisms [42] and often constitute a large fraction of many genomes (45% of the human genome) [42].

TEs are DNA sequences that can change their position inside the genome through a process called transposition (or retrotransposition if the process is RNA mediated). This process can be replicative, i.e. producing another copy that will be inserted in another location in the genome (insertion), or not. The replicative process will invade the host genome, increasing the genome length. TE activity may affect negatively, neutrally and occasionally positively the host fitness, generating a mutual selection owing to the interaction between host and elements living in its genome [43]. With the improvement of genomic sequencing techniques and the increasing of sequencing data, a number of computational tools appeared: 'RepBase', 'Rebase Update', 'Pythia', 'CENSOR', 'RepeatMasker' [44]. These databases and software tools provide information on TE sequences, copy number and their insertion coordinates in the genome.

The analogy between an ecological system and TE has been proposed in [12]. If a copy of a TE is considered as an individual, one TE species comprises closely genetically related TE copies

that share the same interaction with their environment. The community of a genome contains all the copies of TEs irrespective of their subfamilies, families or classes, and it is analogous to the biotic portion of an ecosystem. The abiotic component is composed by the genes and various kinds of noncoding sequences and the intracellular environment.

It is still not clear whether TE community structure is mainly owing to specific element–host interaction rather than stochastic processes; thus, both niche and neutral theories of ecology should and can be taken into account.

Following from the concept of ecological niche, the ‘genomic niche’ for each TE species is inferred from its limiting factors and its genomic traits, which are any biological feature involved in its relationship with the genome. Traits could include preferential site attachment in the genome, strategy of replication, propensity to undergo horizontal transfer (transfer to a new genome out of the germ line). The neutral theory of ecology [45, 46] instead considers a dynamics driven by stochastic forces at the level of individuals, providing an alternative explanation to the establishment of complex TE communities.

Technologies and computational tools

Metabolomics

Metabonomics portrays a well-recognized SM approach involving the study of metabolic responses to stimuli with the aim to understand the dynamics as well as organ-specific biochemical responses [47, 48]. Metabolomics generates multivariate information in cells, tissues and multi-compartmental biological systems, which are a reflection of changes in biological processes [31, 49, 50]. By the global study of low molecular weight compounds (<1500 Da), it provides the characterization of individual metabolic phenotype. Because specific physiological states, gene expression and environmental stressors can cause changes in the steady state of a biological system, monitoring metabolic perturbations provides a unique insight into intra- and extracellular regulatory processes involved in the metabolism [51–53]. To analyze the metabolites, compromising the metabolome, at different concentration and in multiple cellular compartments, different analytical technologies are jointly used. These metabolites include intermediaries of the metabolism, signaling molecules such as hormones and other secondary metabolites [48]. These techniques are mainly based on ¹H nuclear magnetic resonance (NMR) and mass spectrometry coupled gas/high-performance liquid chromatography, with the addition of ultrahigh-performance liquid chromatography systems coupled to mass spectrometry [54]. NMR-based metabolomics provides efficient high-throughput analysis to holistically profile hundreds of metabolites with no a priori selection, while MS is characterized by its increased sensitivity, allowing precise quantification but, owing to increased steps in sample preparation, on a reduced number of samples. The two methods are comprehensively used to generate multivariate information, which are then de-convoluted, by advanced statistical tools, to provide meaningful biological readouts [55, 56]. Metabolomics applications are today directly implemented to large population-based studies to enhance our understanding of the role of genetics, environmental factors and their interactions on individual susceptibility to disease and health [57].

Epigenomics

The inclusion of epigenetic data could play a consistent role in SM applications. The term ‘epigenetics’ defines the interplay

between the genetic background and the environment that, through different molecular mechanisms, eventually produces the observed phenotypes. In particular, epigenetic mechanisms influence the expression of the genes and the penetrance of the different allelic variants, without changing, in term of sequence, the DNA [58]. Among the different epigenetic mechanisms such as micro RNA, proteins ubiquitination, histone methylation and acetylation, one of the most studied one is the methylation of DNA.

DNA methylation occurs in specific genomic sequences, i.e. the CpG dinucleotide motif, where a cytosine is followed by a guanine (Fig. 5). These CpG sites are about 30 million throughout the genome, and there are small genomic regions of about 500–2000 bp that are significantly enriched in CpG sites, which are called ‘CpG islands’, classified as genic CpG islands and non-genic CpG islands, if they map nearby a gene or if they map in a gene desert region accordingly [59].

While the role of the non-genic CpG islands is still not clear, the genic CpG islands are well-recognized functional elements, whose methylation is often correlated with the expression of the nearby gene. Even though this is not a stringent rule, in general, the hypo-methylation of a genic CpG island is correlated with the hyper-expression of the correspondent gene, and vice versa, the hyper-methylation correlates with its hypo-expression.

The state of the art in epigenomic technology is the NGS, which has been successfully applied to the study of the DNA methylation. The field of human aging has gained particularly benefits by these technologies. Indeed, the age-related methylation changes are among one of the major molecular remodeling phenomena that occurs during aging [60–62]. These results strongly support the idea that DNA methylation on the whole is deeply involved in the aging process and its related effects on physiological fitness. This poses DNA methylation as one of the sharpest arrows for the system medicine arch. Nevertheless there are many issues on the use of such data and on how integrate them with the other omics markers.

Next-generation sequencing

The application of NGSs has become a powerful approach for determining DNA and RNA sequences in ‘omics’ studies (e.g. genomics, viromics, metagenomics, transcriptomics and epigenomics). The NGS techniques have been widely used to profile genetic variation of human and environmental samples, as they permit high-resolution and high-throughput detection of DNA and RNA polymorphisms.

The main steps of detecting DNA/RNA polymorphisms, the also called ‘variants’, in NGS data are: quality control check, mapping reads to a reference genome, alignment processing, variant calling and annotation.

The quality control check aims to remove sequence artifacts produced by sequencing platforms. It is performed by trimming or filtering reads according to base quality scores thresholds. The FASTX-Tool kit (http://hannonlab.cshl.edu/fastx_toolkit/) is an example of software that offers a collection of tools for quality control on sequencing data.

The accuracy of the read mapping step has a crucial role in variant detection, as reads that are not correctly aligned can lead to errors in variant and genotype classification. Recently, several fast and memory-efficient software solutions were developed, such as Bowtie2 [63], SOAPv2 [64] and BWA [65]. The manipulation and the processing steps of indexing, sorting and

duplicate removal of alignment files can be done by Picard (<http://broadinstitute.github.io/picard/>) and SAMtools [66].

Previous studies demonstrated that the Phred-scaled quality scores produced by the sequencing platforms may not accurately reflect the true base-calling error rates [67, 68]. As the variant calling algorithms depend on these scores, to assure their accuracy, a recalibration process can be performed by SOAPsnp [68] and GATK [69].

Determining polymorphisms in sequencing data, also referred as ‘variant calling’, permits the classification of allelic genotypes and, depending on the experimental design, also permits the discovery of molecular events such as somatic mutations, loss of heterozygosity, mRNA editing and clonal evolution of cancer samples. This kind of information is essential to differentiate between ‘normal’ and tumoral tissues and can be used for the construction of biomarker signatures useful to provide computational tools for early diagnosis of cancer risk. The variant calling can be performed by general-purpose software products, as GATK and SAMtools, or by software solutions designed to specific purposes, like VarScan2 [70], SomaticSniper [71], JointSNVMix [72], Strelka [73] and MuTect [74], that were developed for the joint analysis of matched normal-cancer samples.

The final step is the variant annotation, and so variants can be prioritized according to their biological meaning and/or disease-causing potential. The Annovar [75] and SnpEff [76] are examples of software designs that perform gene-based annotation of human variants, which take in account information also present in dbSNP and 1000 Genomes databases.

The choice of tools and parameters for sequencing data analysis may have great impact on the final results and it will vary according to the type of sample, sequencing platform and experimental design.

NGS methods are widely used in metagenomics and are based on the taxonomic value of 16S rRNA for the microbial classification. Usually 16S rRNA are massively sequenced by various methods (e.g. 454, ion torrent) and after that, the main step of 16S rRNA sequencing data processing is indeed clustering. Clustering means grouping sequences according to some similarity criterion and it is exploited by two fundamental analysis: taxonomic classification of bacteria in the sample and ecological description of such population.

The two widely used approaches put sequences into bins based on either their similarity to reference sequences (i.e. phylotyping by reference clustering) or their similarity to other sequences in the community [i.e. Operational Taxonomic Units (OTUs)] [77].

Reference clustering is of course helpful if one’s aim is taxonomic classification, but it implies a loss of information that makes it almost useless if the purpose is to give an ecological description of the microbiota community.

The distribution of OTUs (RSA, see Figure 4) is an important component of biodiversity. To assess the microbiota RSA from 16S rRNA sequencing data, the process is to cluster sequences with a de novo algorithm, thus producing the so-called OTUs. Because 16S rRNA gene is a highly conserved sequence of the genome, sequences that belong to the same OTU will be phylogenetically closed and could be used as ‘Species’ to compute the RSA.

Many different algorithms have been suggested to carry out de novo clustering: BLAST [78], MOTHUR [79], QIIME [80], CD-HIT [81] and UCLUST [82]. In particular, UCLUST is a greedy algorithm that uses less memory and is even faster than CD-HIT [82, 83]. It starts with an empty database in memory and then reads

the sequences in input order. Sequences with a similarity greater than 97% are typically assigned to the same species, those with similarity >95% to the same genus, and those with >80% to the same phylum. OTUs defined by a certain similarity threshold represent therefore bacteria ‘species’ at a particular phylogenetic level. It follows that starting from OTU’s abundances, one can easily build the RSA distribution by plotting the number of species (y-axis) that have a certain number of individuals, i.e. sequences, (x-axis), and have a measure of the microbiota biodiversity. Because of the RSA long tail, it is usual to consider the logarithm to base two of the x-axis, obtaining the so-called Preston plot (Figure 4).

A remarkable result is that (see Figure 4) the RSA curves obtained from microbial and TE follow the same statistical distribution. In particular, a population dynamics explanation has been given in [45] and [46], by a stochastic neutral model in which species interaction is neglected and the resulting RSA is described by a Negative Binomial Distribution. The parameters of this distribution are particularly sensitive to variations such as dietary regimes (e.g. the amount of proteins, fatty acids and fibers), pathological conditions (e.g. obesity, MS and type 2 diabetes) and aging. The characterization (even in the multidimensional case) of these parameters provides a further and advanced method aimed to classify different health states.

Multilayer networks: a computational tool for multi-scale integration and diffusion

Network theory is a major branch of the so-called complexity science: it investigates the global topology and structural patterns of the interactions among the constituents of social groups, infrastructures, brain and biological networks [15, 84, 85]. Networks are the most natural way to model many types of biological knowledge at different levels: from single-cell protein interaction, transcriptional regulation and metabolic networks, up to neuronal, IS or ecological networks involving multiple cells or organisms.

Network approaches can be useful, in a Personalized Medicine perspective [3], for the following purposes:

- Community detection (e.g. identification of functional modules or patient stratification) [86, 87].
- Centrality measures for ranking network elements (nodes, edges), from classical measures like Betweenness Centrality, to more recent ones, also suited for dense and weighted networks [88, 89].
- Characterization of network ensembles as null models, to identify nonrandom patterns in real networks and for multilevel network analysis embedding a priori biological information [90–92].

The omics era has converged to the era of ‘big data’ for all types of biological information: genomics, proteomics, interactomics, cellomics, organomics, *in vitro* and *in vivo* imaging [3], and it has generated a new approach to medicine, i.e. SM. The definition of SM is deeply related to complex networks: it involves a systemic view of the organism where the various building elements are considered in their interplay [16, 93, 94]. A paradigmatic example is the networks of human diseases. The diseasesome [95] is a bipartite network that connects human diseases (disease phenome) and human genes (disease genome). The strong connectivity of molecular systems implies that a specific dysfunction of an element propagates throughout the network of interactions and affects the activity of other components. Therefore, different dysfunctions can lead to the same

effect through different pathways. The concept of ‘disease module’ has been proposed to indicate a group of network components whose disruption leads to a particular disease phenotype.

Single-target therapies may fail because they are not contemplating the underlying network characteristics during target identification [96]. On the other side, a drug used for a disease could reveal itself of valuable use also for other diseases strongly connected to the original target node. Two examples of network-based disease targeting are the ‘central hit strategy’, that selectively focuses on central nodes/edges in the biological network and the ‘network influence strategy’, in which several neighbors of central nodes/edges are targeted instead [94]. These approaches require detailed knowledge of the underlying networks and of its dynamics, as for example by controllability approaches, i.e. the study of how specific nodes can shift the dynamical status of a network [96–99].

Each patient status can be described by several networks, e.g. related to its genomic, proteomic or metabolic profile. When gene expression data are available, they can be mapped onto protein–protein networks, thus integrating general biological knowledge and sample-specific information at multiple levels (from the whole network to a single gene or pathway) [100].

Further, each patient can be summarized by a vector of network observables, like betweenness centrality, spectral centrality [88], community labels [86] or single-node entropy measures [100], and suitable clustering/discrimination algorithms can help to stratify or classify such samples. A possible vector is related to the concept of diffusion of information throughout a network. This measure has been applied to the stratification of tumor mutations [101], the association of genes and protein complexes with diseases [102], the identification of biomarkers in genome-wide studies [103, 104] and the relation between viral (Epstein Barr Virus (EBV), Human Papilloma Virus (HPV) and Hepatitis C Virus (HCV)) targets and the corresponding cellular response [105, 106]. Diffusion algorithms simulate the diffusion of information from a subset of vertexes to all the others. At the end of the process, the amount of information per vertex depends both on the initial distribution of information and on network topology; therefore, it permits to rank all vertexes in relation to their network proximity to the subset of vertexes that carry the initial information.

So far, all the different networks for the same sample have been investigated separately. Recently, framework of ‘multilayer networks’ has been introduced [16, 107–109]. SM seems naturally embedded in a multilayer network.

Multilayer networks might represent the tools to successfully combine the interactions between the different constituents of the cell, and the answer to the purpose of personalized medicine to define a ‘quantified self’ or for assessing global wellness [3]. For example, a multilayer network provides novel approaches for the combined analysis of samples in different states [109].

Integrative multi-omics approaches offer a more comprehensive picture, especially when embedded in a priori biological knowledge [110], and different layers of disease-related information can be analyzed altogether [111]. Some multilayer integrative approaches, like iCluster [112] and MDI (Multiple Dataset Integration) [113], can be applied to several types of data sets, while others, such as Camelot [114] and CNAmets [115], were developed for specific combinations of omics. The integration strategies can be sequential or simultaneous: sequential approaches refine the results of one layer using further layers; conversely, simultaneous approaches jointly analyze all layers without imposing an a priori order.

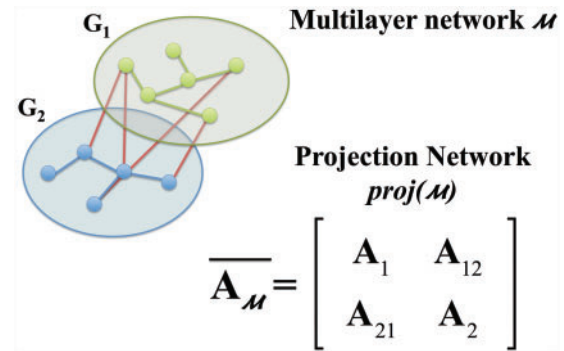


Figure 6. A representation of a generic multilayer network composed by two graphs: G_1 and G_2 . The interlayer connections are in red, while the intralayer connections are in green for graph G_1 and in blue for graph G_2 . The adjacency matrix of the related projection network $\text{proj}(\mathcal{M})$ is displayed in the lower-right corner. A colour version of this figure is available at BIB online: <http://bib.oxfordjournals.org>.

Formally speaking [16], a multilayer graph \mathcal{M} is described by a set of M graphs $\{G_x\}$ (called layers) and by a set of connections between nodes of different G_x and G_β (‘interlayer connections’, see Figure 6, red edges). ‘Intralayer connections’ connect nodes in the same layer (Figure 6, green and blue edges). We denote with $\mathbf{A}^x = \{a_{ij}^x\} \in \mathbb{R}^{N_x \times N_x}$ the adjacency matrix of the network layer G_x and by $\mathbf{A}^{x\beta} = \{a_{ij}^{x\beta}\} \in \mathbb{R}^{N_x \times N_\beta}$ the adjacency matrix representing interlayer connections between nodes of different G_x and G_β . The adjacency matrix of the projection network is indicated as $\overline{\mathbf{A}}_{\mathcal{M}}$ (shown in Figure 6).

Multilayer networks can be distinguished in multiplex networks [116–119] and interacting networks of networks [120, 121]. In a multiplex network, the same set of nodes has different types of interactions in each layer. In the latter, nodes of different layers represent different elements of the system, as in the cell, where metabolites, proteins and transcription factors are distinct biological entities [109]. These systems show novel critical phenomena and dynamical processes [122–124]. Multiplex features like degree correlation [125] determine whether a hub in one layer is also a hub in another layer, while the overlap quantifies how much two nodes of the network interact in several networks at the same time. Recent application to social or biological weighted networks [108, 109] show how novel and relevant information can be uncovered only by exploiting the multiplex nature of such systems. In a typical case-control design, e.g. cancer versus healthy samples, co-expression networks (based on sample omics measurements) are naturally embedded in a case-control duplex [109]. This duplex naturally stratifies in the backbone of interactions shared by both layers, highlighting layer-specific relations and behaviors.

SM approaches can be embedded in network of networks [3] or in multiplex networks (Figure 7), according to the model assumptions and to the available observables. The single-layer measures must be modified to take into account the multilayered structure, such as for the definition of shortest paths [126]. Centrality measures [127] and Network entropy [100, 107, 109] can be redefined onto such structures, extending their meaning over different types of nodes or relationships. A still unexplored field concerns the use of information in some layers as priors for network-based statistical analysis of a single layer. This can be exploited for network reconstruction or for multidimensional clustering [128], with the features of one layer used as constraints for the other layers.

A multiplex network approach has been used to find recurrent subgraphs in co-expression networks from biological data

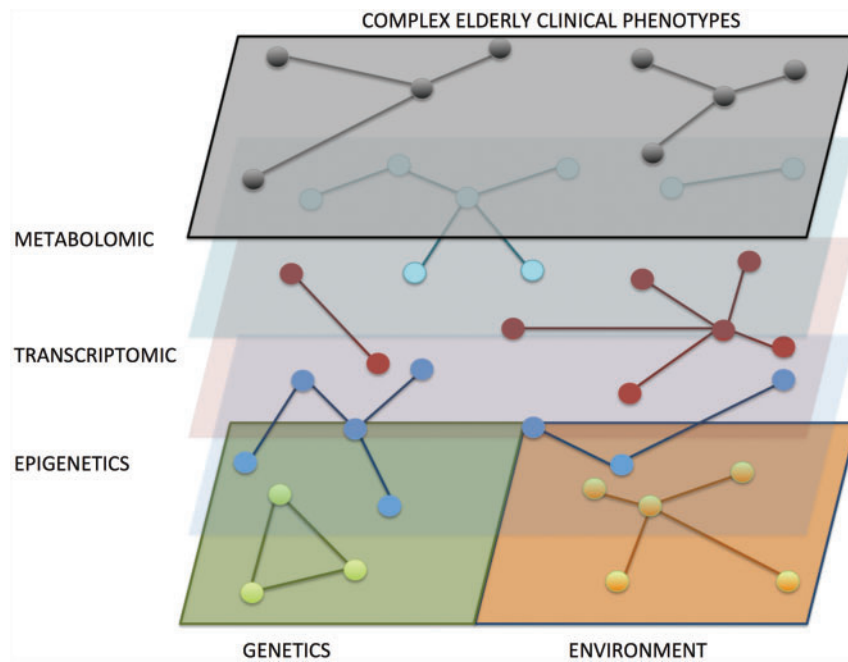


Figure 7. A possible multiplex architecture for SM: in a multiplex network, the same set of nodes (patients) has different types of interactions in each layer. Each layer corresponds to a given omics measurement, and a link might be a similarity measure between two people. We represent only few layers for the sake of simplicity. The bottom layer is divided into Genetics and Environment that may have a different role in the causation of a given phenotypical trait (e.g. clusters). A colour version of this figure is available at BIB online: <http://bib.oxfordjournals.org>.

[129]. These subgraphs can be associated with functional modules, protein complexes or transcriptional modules, for particular phenotypic conditions, confirming how phenotypes are related to complex structures of the genetic network. In the same context, sub-network extraction based on multi-objective optimization of several criteria permits the identification of modules enriched in multi-omics information [130].

Last but not least, multiplex networks are the ideal formalism for time-evolving networks [131]: if the same measurements are available at different stages (e.g. at disease onset, during and after therapy), neglecting time-dependence of the links compromises, from the point of view of the personalized medicine, the correct reconstruction of patient's history and disease progression.

Conclusion and perspectives

In this article we discussed the hypothesis that inflammation/inflammaging can be a unifying mechanism for SM. The reasons are rooted in: (i) the multi-scale character of inflammaging, that starts as a local phenomenon and subsequently propagates at a systemic level; (ii) the involvement of the microenvironment as an active medium for exchanging signals between different layers of biological complexity; (iii) the possibility to use ecological modeling to describe and predict the behavior of systems such as the GM and the Mobilome; (iv) the natural embedding of SM in a multilayer network formalism, capable to integrate multi-omics data.

Future challenges will investigate ecological niche and association dynamics of microbiomes by using 3D cartography of human body [132]. One important direction will be the study of the impact of human microbiome distribution on immunity and on the tissue-specific epigenomic marks of disease- and trait-associated genetic variants (see [133], for a large collections of human epigenomes). This could be done for different

populations, human mobility groups in routine [133] and shift conditions. This integrated analysis will give ground for a better understanding of the evolution of comorbidities and multi-morbidities, particularly in presence of infectious diseases and for finding appropriate therapies [134, 135].

We strongly believe that the combination of these concepts and tools will be the bases for a SM focused on the comprehension of patho-physiological processes such as aging and age-related pathologies, including cancer, and they will pave the way toward a better understanding of the differences between health and disease states.

Key Points

- Inflammation is a unifying Biomedical hypothesis particularly attractive for Systems Medicine as it is multi-scale, multi-organs and propagating among multiple spatial and temporal scales.
- Nowadays sequencing technologies are among the most powerful methods to achieve omics information from biological systems (Metagenome, Virome, Mobilome).
- All biological systems are communicating, so the Communicome is crucial, as well as its metabolomic characterization.
- An emerging strategy for the description and prediction of interconnected biological systems, including tissues and organs, is to model them as ecosystems, remarking the role of the environment in the propagation process.
- The human body, within the Systems Medicine vision, can be mapped onto a multilayer network, capable to model and quantify the endogenous and exogenous interactions.

Funding

All the authors acknowledge support by the Italian Ministry of Education and Research through the Flagship (PB05) InterOmics and EU projects FibeBiotics (GA: 289517), Mission-T2D (GA: 600803) and Methods for Integrated analysis of multiple Omics data sets (MIMOmics) (GA: 305280).

Reference

1. Wolkenhauer O, Auffray C, Brass O, et al. Enabling multiscale modeling in systems medicine. *Genome Med* 2014;6:21.
2. Zhang H, Gustafsson M, Nestor C, et al. Targeted omics and systems medicine: personalising care. *Lancet Respir Med* 2014;2:785–7.
3. Hood L, Flores M. A personal view on systems medicine and the emergence of proactive P4 medicine: predictive, preventive, personalized and participatory. *N Biotechnol* 2012;29:613–24.
4. Franceschi C, Bonafè M, Valensin S, et al. Inflamm-aging: an evolutionary perspective on immunosenescence. *Ann NY Acad Sci* 2006;908:244–54.
5. Franceschi C, Campisi J. Chronic inflammation (inflammaging) and its potential contribution to age-associated diseases. *J Gerontol A Biol Sci Med Sci* 2014;69 (Suppl 1):S4–9.
6. Franceschi C, Capri M, Monti D, et al. Inflammaging and anti-inflammaging: a systemic perspective on aging and longevity emerged from studies in humans. *Mech Ageing Dev* 2007;128:92–105.
7. Auffray C, Chen Z, Hood L. Systems medicine: the future of medical genomics and healthcare. *Genome Med* 2009;1:2.
8. Ray S, Britschgi M, Herbert C, et al. Classification and prediction of clinical Alzheimer's diagnosis based on plasma signaling proteins. *Nat Med* 2007;13:1359–62.
9. Belyakov OV, Malcolmson AM, Folkard M, et al. Direct evidence for a bystander effect of ionizing radiation in primary human fibroblasts. *Br J Cancer* 2001;84:674–9.
10. Nelson G, Wordsworth J, Wang C, et al. A senescent cell bystander effect: senescence-induced senescence. *Aging Cell* 2012;11:345–9.
11. Bernard C. *An Introduction to the Study of Experimental Medicine*, translated by Greene HC. Mineola, NY: Dover Publications, 1957.
12. Venner S, Feschotte C, Biéumont C. Dynamics of transposable elements: towards a community ecology of the genome. *Trends Genet* 2009;25:317–23.
13. Virgin HW. The virome in mammalian physiology and disease. *Cell* 2014;157:142–50.
14. Koonin EV, Dolja VV, Krupovic M. Origins and evolution of viruses of eukaryotes: the ultimate modularity. *Virology* 2015;479–80:2–25.
15. Albert R, Barabási AL. Statistical mechanics of complex networks. *Rev Mod Phys* 2002;74:47.
16. Boccaletti S, Bianconi G, Criado R, et al. The structure and dynamics of multilayer networks. *Phys Rep* 2014;544:1–122.
17. Kivela M, Arenas A, Barthelemy M, et al. Multilayer networks. *J Complex Networks* 2014;2:203–71.
18. Biagi E, Nylund L, Candela M, et al. Through ageing, and beyond: gut microbiota and inflammatory status in seniors and centenarians. *PLoS One* 2010;5:e10667.
19. Sansoni P, Vescovini R, Fagnoni FF, et al. New advances in CMV and immunosenescence. *Exp Gerontol* 2014;55:54–62.
20. Dall'Olio F, Vanhooren V, Chen CC, et al. N-glycomic biomarkers of biological aging and longevity: a link with inflammaging. *Ageing Res Rev* 2013;12:685–98.
21. Olivieri F, Rippo MR, Monsurrò V, et al. MicroRNAs linking inflamm-aging, cellular senescence and cancer. *Ageing Res Rev* 2013;12:1056–68.
22. Coppé JP, Desprez PY, Krtolica A, et al. The senescence-associated secretory phenotype: the dark side of tumor suppression. *Annu Rev Pathol* 2010;5:99–118.
23. Jurk D, Wilson C, Passos JF, et al. Chronic inflammation induces telomere dysfunction and accelerates ageing in mice. *Nat Commun* 2014;2:4172.
24. Vitale G, Salvioli S, Franceschi C. Oxidative stress and the ageing endocrine system. *Nat Rev Endocrinol* 2013;9:228–40.
25. Tchkonja T, Zhu Y, van Deursen J, et al. Cellular senescence and the senescent secretory phenotype: therapeutic opportunities. *J Clin Invest* 2013;123:966–72.
26. Berendsen A, Santoro A, Pini E, et al. A parallel randomized trial on the effect of a healthful diet on inflammaging and its consequences in European elderly people: design of the NU-AGE dietary intervention study. *Mech Ageing Dev* 2013;134:523–30.
27. Ostan R, Lanzarini C, Pini E, et al. Inflammaging and cancer: a challenge for the mediterranean diet. *Nutrients* 2015;7:2589–621.
28. Focà A, Liberto MC, Quirino A, et al. Gut inflammation and immunity: what is the role of the human gut virome? *Mediators Inflamm* 2015;2015:26032.
29. Peterson J, Garges S, Giovanni M, et al. The NIH human microbiome project. *Genome Res* 2009;19:2317–23.
30. Handelsman J, Rondon MR, Brady SF, et al. Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products. *Chem Biol* 1998;5:R245–9.
31. Nicholson JK, Wilson ID. Opinion: understanding 'global' systems biology: metabonomics and the continuum of metabolism. *Nat. Rev. Drug Discov* 2003;2:668–76.
32. Hooper LV, Littman DR, Macpherson AJ. Interactions between the microbiota and the immune system. *Science* 2012;336:1268–73.
33. Blumberg R, Powrie F. Microbiota, disease, and back to health: a metastable journey. *Sci Transl Med* 2012;4:137rv7.
34. Goldsmith JR, Sartor RB. The role of diet on intestinal microbiota metabolism: downstream impacts on host immune function and health, and therapeutic implications. *J Gastroenterol* 2014;49:785–98.
35. Cani PD, Delzenne NM. The role of the gut microbiota in energy metabolism and metabolic disease. *Curr Pharm Des* 2009;15:1546–58.
36. Turnbaugh PJ, Ley RE, Mahowald MA, et al. An obesity-associated gut microbiome with increased capacity for energy harvest. *Nature* 2006;444:1027–31.
37. Claesson MJ, Jeffery IB, Conde S, et al. Gut microbiota composition correlates with diet and health in the elderly. *Nature* 2012;488:178–84.
38. Ibrahim K, Jusoff K, Md Isa NN. Biological diversity assessment of Tok Bali mangrove forest, Kelantan, Malaysia. *WSEAS Trans Environ Dev* 2007; 37–44. ISSN 1790–5079.
39. Preston FW. The commonness, and rarity, of species. *Ecology* 1948;29:254.
40. Hubbell SP. *The unified neutral theory of biodiversity and biogeography (MPB-32)*. 2011.
41. Chesson P. Mechanisms of maintenance of species diversity. *Annu Rev Ecol Syst* 2000;31:343–36.

42. Muñoz-López M, García-Pérez JL. DNA transposons: nature and applications in genomics. *Curr Genomics* 2010;11:115–28.
43. Rebollo R, Romanish MT, Mager DL. Transposable elements: an abundant and natural source of regulatory sequences for host genes. *Annu Rev Genet* 2012;46:21–42.
44. Jurka J. Repbase update: a database and an electronic journal of repetitive elements. *Trends Genet* 2000;16:418–20.
45. Azaele S, Pigolotti S, Banavar JR, et al. Dynamical evolution of ecosystems. *Nature* 2006;444:926–8.
46. Volkov I, Banavar JR, Hubbell SP, et al. Patterns of relative species abundance in rainforests and coral reefs. *Nature* 2007;450:45–9.
47. Nicholson JK, Lindon JC, Holmes E. 'Metabonomics': understanding the metabolic responses of living systems to pathophysiological stimuli via multivariate statistical analysis of biological NMR spectroscopic data. *Xenobiotica* 1999;29:1181–9.
48. Collino S, Martin FPJ, Kochhar S, et al. Nutritional metabonomics: an approach to promote personalized health and wellness. *Chim Int J Chem* 2011;65:396–9.
49. Nicholson JK, Lindon JC. Systems biology: metabonomics. *Nature* 2008;455:1054–6.
50. Holmes E, Wilson ID, Nicholson JK. Metabolic phenotyping in health and disease. *Cell* 2008;134:714–17.
51. Keurentjes JJB. Genetical metabolomics: closing in on phenotypes. *Curr Opin Plant Biol* 2009;12:223–30.
52. Saito K, Matsuda F. Metabolomics for functional genomics, systems biology, and biotechnology. *Annu Rev Plant Biol* 2010;61:463–89.
53. Weckwerth W. Metabolomics in systems biology. *Annu Rev Plant Biol* 2003;54:669–89.
54. Martin FPJ, Collino S, Rezzi S. 1H NMR-based metabonomic applications to decipher gut microbial metabolic influence on mammalian health. *Magn Reson Chem* 2011;49 (Suppl 1):S47–54.
55. Martin FPJ, Montoliu I, Kochhar S, et al. Chemometric strategy for modeling metabolic biological space along the gastrointestinal tract and assessing microbial influences. *Anal Chem* 2010;82:9803–11.
56. Montoliu I, Martin FPJ, Collino S, et al. Multivariate modeling strategy for intercompartmental analysis of tissue and plasma 1H NMR spectrotypes. *J Proteome Res* 2009;8:2397–406.
57. Collino S, Montoliu I, Martin FPJ, et al. Correction: metabolic signatures of extreme longevity in northern Italian centenarians reveal a complex remodeling of lipids, amino acids, and gut microbiota metabolism. *PLoS One* 2013;8. doi:10.1371/annotation/5fb9fa6f-4889-4407-8430-6dfc7ec
58. Bird A. DNA methylation patterns and epigenetic memory. *Genes Dev* 2002;16:6–21.
59. Bacalini MG, Boattini A, Gentilini D, et al. A meta-analysis on age-associated changes in blood DNA methylation: results from an original analysis pipeline for Infinium 450k data. *Aging (Albany, NY)* 2015;7:97–109.
60. Hannum G, Guinney J, Zhao L, et al. Genome-wide methylation profiles reveal quantitative views of human aging rates. *Mol Cell* 2013;49(2):359–67.
61. Bacalini MG, Gentilini D, Boattini A, et al. Identification of a DNA methylation signature in blood cells from persons with Down Syndrome. *Aging (Albany, NY)* 2015;7:82–96.
62. Garagnani P, Bacalini M, Pirazzini C. Methylation of ELOVL2 gene as a new epigenetic marker of age. *Aging Cell* 2012;11(6):1132–4.
63. Langmead B, Salzberg S. Fast gapped-read alignment with Bowtie 2. *Nat Methods* 2012;9(4):357–9.
64. Li R, Yu C, Li Y, et al. SOAP2: an improved ultrafast tool for short read alignment. *Bioinformatics* 2009;25:1966–7.
65. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 2009;25:1754–60.
66. Li H, Handsaker B, Wysoker A, et al. The sequence alignment/map format and SAMtools. *Bioinformatics* 2009;25:2078–9.
67. Brockman W, Alvarez P, Young S, et al. Quality scores and SNP detection in sequencing-by-synthesis systems. *Genome Res* 2008;18:763–70.
68. Li R, Li Y, Fang X, et al. SNP detection for massively parallel whole-genome resequencing. *Genome Res* 2009;19:1124–32.
69. DePristo MA, Banks E, Poplin R, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet* 2011;43:491–8.
70. Koboldt DC, Zhang Q, Larson DE, et al. VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Res* 2012;22:568–76.
71. Larson DE, Harris CC, Chen K, et al. SomaticSniper: identification of somatic point mutations in whole genome sequencing data. *Bioinformatics* 2012;28:311–17.
72. Roth A, Ding J, Morin R, et al. JointSNVMix: a probabilistic model for accurate detection of somatic mutations in normal/tumour paired next-generation sequencing data. *Bioinformatics* 2012;28:907–13.
73. Saunders CT, Wong WSW, Swamy S, et al. Strelka: accurate somatic small-variant calling from sequenced tumor-normal sample pairs. *Bioinformatics* 2012;28:1811–17.
74. Cibulskis K, Lawrence MS, Carter SL, et al. Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nat Biotechnol* 2013;31:213–19.
75. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res* 2010;38(16):e164.
76. Cingolani P, Platts A, Wang LL, et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly (Austin)* 2012;6:80–92.
77. Schloss PD, Westcott SL. Assessing and improving methods used in operational taxonomic unit-based approaches for 16S rRNA gene sequence analysis. *Appl Environ Microbiol* 2011;77:3219–26.
78. Altschul SF, Madden TL, Schäffer AA, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 1997;25:3389–402.
79. Schloss PD, Westcott SL, Ryabin T, et al. Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl Environ Microbiol* 2009;75:7537–41.
80. Caporaso JG, Kuczynski J, Stombaugh J, et al. QIIME allows analysis of high-throughput community sequencing data. *Nat Methods* 2010;7:335–6.
81. Li W, Godzik A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* 2006;22:1658–9.
82. Edgar RC. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* 2010;26:2460–1.
83. Costello EK, Lauber CL, Hamady M, et al. Bacterial community variation in human body habitats across space and time. *Science* 2009;326:1694–7.

84. Newman MEJ. Mixing patterns in networks. *Phys Rev E Stat Nonlin Soft Matter Phys* 2003;**67**:026126.
85. Boccaletti S, Latora V, Moreno Y, et al. Complex networks: structure and dynamics. *Phys Rep* 2006;**424**:175–308.
86. Girvan M, Newman MEJ. Community structure in social and biological networks. *Proc Natl Acad Sci USA* 2002;**99**:7821–6.
87. Fortunato S. Community detection in graphs. *Phys Rep* 2010;**486**:75–174.
88. Pauls SD, Remondini D. Measures of centrality based on the spectrum of the Laplacian. *Phys Rev E Stat Nonlinear Soft Matter Phys* 2012;**85**:066127.
89. Grady D, Thiemann C, Brockmann D. Robust classification of salient links in complex networks. *Nat Commun* 2012;**3**:864.
90. Park J, Newman MEJ. Statistical mechanics of networks. *Phys Rev E* 2004;**70**:66117.
91. Bianconi G. Entropy of network ensembles. *Phys Rev E Stat Nonlinear Soft Matter Phys* 2009;**79**:1–10.
92. Squartini T, Garlaschelli D. Analytical maximum-likelihood method to detect patterns in real networks. *New J Phys* 2011;**13**:083001.
93. Kolodkin A, Boogerd FC, Plant N, et al. Emergence of the silicon human and network targeting drugs. *Eur J Pharm Sci* 2012;**46**:190–7.
94. Csérmely P, Korcsmáros T, Kiss HJM, et al. Structure and dynamics of molecular networks: a novel paradigm of drug discovery: a comprehensive review. *Pharmacol Ther* 2013;**138**:333–408.
95. Goh KI, Cusick ME, Valle D, et al. The human disease network. *Proc Natl Acad Sci USA* 2007;**104**:8685–90.
96. Luni C, Shoemaker JE, Sanft KR, et al. Confidence from uncertainty—a multi-target drug screening method from robust control theory. *BMC Syst Biol* 2010;**4**:161.
97. Liu YY, Slotine JJ, Barabási AL. Controllability of complex networks (SI). *Nature* 2011;**473**:167–73.
98. Xiong H, Choe Y. Dynamical pathway analysis. *BMC Syst Biol* 2008;**2**:9.
99. Menichetti G, Dall'Asta L, Bianconi G. Network controllability is determined by the density of low in-degree and out-degree nodes. *Phys Rev Lett* 2014;**113**:078701.
100. Menichetti G, Bianconi G, Castellani G, et al. Multiscale characterization of aging and cancer progression by a novel Network Entropy measure. *Mol BioSyst* 2015.
101. Hofree M, Shen J, Carter H, et al. Network-based stratification of tumor mutations. *Nat Methods* 2013;**10**:1108–15.
102. Vanunu O, Magger O, Ruppin E, et al. Associating genes and protein complexes with disease via network propagation. *PLoS Comput Biol* 2010;**6**:e1000641.
103. Stokes ME, Barmada MM, Kamboh MI, et al. The application of network label propagation to rank biomarkers in genome-wide Alzheimer's data. *BMC Genomics* 2014;**15**:282.
104. Qian Y, Besenbacher S, Mailund T, et al. Identifying disease associated genes by network propagation. *BMC Syst Biol* 2014;**8**(Suppl 1):S6.
105. Gulbahce N, Yan H, Dricot A, et al. Viral perturbations of host networks reflect disease etiology. *PLoS Comput Biol* 2012;**8**:e1002531.
106. Mosca E, Alfieri R, Milanesi L. Diffusion of information throughout the host interactome reveals gene expression variations in network proximity to target proteins of hepatitis C virus. *PLoS One* 2014;**9**:e113660.
107. Bianconi G. Statistical mechanics of multiplex ensembles: entropy and overlap. *Phys Rev E* 2013;**87**:062806.
108. Menichetti G, Remondini D, Panzarasa P, et al. Weighted multiplex networks. *PLoS One* 2014;**9**:e97857.
109. Menichetti G, Remondini D, Bianconi G. Correlations between weights and overlap in ensembles of weighted multiplex networks. *Phys Rev E* 2014;**90**:062817.
110. Kristensen VN, Lingjærde OC, Russnes HG, et al. Principles and methods of integrative genomic analyses in cancer. *Nat Rev Cancer* 2014;**14**:299–313.
111. Gustafsson M, Nestor CE, Zhang H, et al. Modules, networks and systems medicine for understanding disease and aiding diagnosis. *Genome Med* 2014;**6**:82.
112. Shen R, Olshen A, Ladanyi M. Integrative clustering of multiple genomic data types using a joint latent variable model with application to breast and lung cancer subtype analysis. *Bioinformatics* 2009;**25**:2906–12.
113. Kirk P, Griffin JE, Savage RS, et al. Bayesian correlated clustering to integrate multiple datasets. *Bioinformatics* 2012;**28**:3290–7.
114. Chen BJ, Causton HC, Mancenido D, et al. Harnessing gene expression to identify the genetic basis of drug resistance. *Mol Syst Biol* 2009;**5**:310.
115. Louhimo R, Hautaniemi S. CNAmet: an R package for integrating copy number, methylation and expression data. *Bioinformatics* 2011;**27**:887–8.
116. Cardillo A, Gómez-Gardeñes J, Zanin M, et al. Emergence of network features from multiplexity. *Sci Rep* 2013;**3**:1344.
117. Donges JF, Zou Y, Marwan N, et al. Complex networks in climate dynamics. *Eur Phys J Spec Top* 2009;**174**:157–79.
118. Nicosia V, Latora V. Measuring and modelling correlations in multiplex networks. *arXiv Prepr* 2014;**21**:1–21.
119. Szell M, Lambiotte R, Thurner S. Multirelational organization of large-scale social networks in an online world. *Proc Natl Acad Sci* 2010;**107**:13636–41.
120. Gao J, Buldyrev SV, Stanley HE, et al. Networks formed from interdependent networks. *Nat Phys* 2011;**8**:40–8.
121. Bianconi G, Dorogovtsev SN. Multiple percolation transitions in a configuration model of a network of networks. *Phys Rev E* 2014;**89**:062814.
122. Buldyrev SV, Parshani R, Paul G, et al. Catastrophic cascade of failures in interdependent networks. *Nature* 2010;**464**:1025–8.
123. Baxter GJ, Dorogovtsev SN, Goltsev AV, et al. Avalanche collapse of interdependent networks. *Phys Rev Lett* 2012;**109**:248701.
124. Brummitt CD, D'Souza RM, Leicht EA. Suppressing cascades of load in interdependent networks. *Proc Natl Acad Sci USA* 2012;**109**:E680–9.
125. Min B, Yi S Do, Lee K-M, et al. Network robustness of multiplex networks with interlayer degree correlations. *Phys Rev E* 2014;**89**:042811.
126. Battiston F, Nicosia V, Latora V. Structural measures for multiplex networks. *Phys Rev E* 2014;**89**:032804.
127. De Domenico M, Solé-Ribalta A, Omodei E, et al. Centrality in interconnected multilayer networks. *arXiv Prepr* 2013;**12**.
128. Dong X, Frossard P, Vandergheynst P, et al. *IEEE Xplore - Clustering With Multi-Layer Graphs: A Spectral Perspective*. pp. 1–12.
129. Li W, Liu CC, Zhang T, et al. Integrative analysis of many weighted co-expression networks using tensor computation. *PLoS Comput Biol* 2011;**7**:e1001106, 1–13.

130. Mosca E, Milanesi L. Network-based analysis of omics with multi-objective optimization. *Mol Biosyst* 2013;**9**(12):2971–80.
131. Mucha PJ, Richardson T, Macon K, et al. Community structure in time-dependent, multiscale, and multiplex networks. *Science* 2010;**328**:876–8.
132. Bouslimani A, Porto C, Rath CM, et al. Molecular cartography of the human skin surface in 3D. *Proc Natl Acad Sci USA* 2015;**112**:E2120–9.
133. Consortium RE, Kundaje A, Meuleman W, et al. Integrative analysis of 111 reference human epigenomes. *Nature* 2015;**518**:317–30.
134. Lax S, Nagler C, Gilbert J. Our interface with the built environment: immunity and the indoor microbiota. *Trends Immunol* 2015;**36**(3):121–3.
135. Moni M, Liò P. Network-based analysis of comorbidities risk during an infection: SARS and HIV case studies. *BMC Bioinformatics* 2014;**15**:333.